Regresia logistica

Regresia logistica este un algoritm ce rezolva problema clasificarii (variabilele independente sunt continue, iar cele dependente categorice), folosind functia logistica, o functie sigmoida, pentru a incadra rezultatele in intervalul (0,1):

σ (t) = , σ : ℝ→(0,1)

si are scopul de a gasi un model pentru determinarea valorii clasei careia ii apartin datele de intrare, cu o eroare cat mai mica.

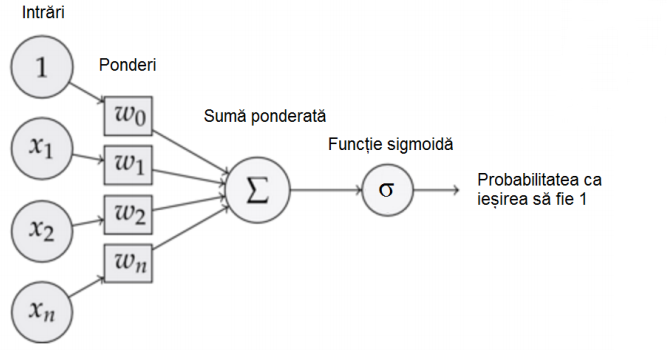


Fig. 1: Regresie logistica

Setul de date de training este format din n puncte de forma:

1, y1), 2, y2),3, y3), ...,n, yn),

unde x = (X1 , X2 , …, Xn ) un vector de intrare

si y ϵ {0,1} .

Presupunem o relatie liniara intre atributele setului de date si ponderile acestora astfel:

t = w0+X1w1+...+Xnwn

Modelul de regresie logistica va fi :

p = ,

unde p este probabilitatea ca x sa aiba valoarea y = 1 asociata.